STATISTICALLY DRIVEN SENTENCE REALIZING

METHOD AND APPARATUS

ABSTRACT OF THE DISCLOSURE

A method of, and system for, generating a sentence from a semantic representation maps the semantic representation to an unordered set of syntactic nodes. Simplified generation grammar rules and statistical goodness measure values from a corresponding analysis grammar are then used to create a tree structure to order the syntactic nodes. The sentence is then generated from the tree structure. The generation grammar is a simplified (context free) version of a corresponding full (context sensitive) analysis grammar. In the generation grammar, conditions on each rule are ignored except those directly related to the semantic representation. The statistical goodness measure values, which are calculated through an analysis training phase in which a corpus of example sentences is processed using the full analysis grammar, are used to guide the generation choice to prefer substructures most commonly found in a particular syntactic/semantic context during analysis.